TEMPTATIONS OF THE FLESCH

G. Harry McLaughlin

August, 1969

Occasional Paper No. 2

NGL 33-022-090

<u>Temptations of the Flesch</u>

G. Harry McLaughlin

Communications Research Center
Syracuse University

There is no such thing as a valid readability formula. Readability is
generally taken to mean that quality of written material which induces a reader
to go on reading. Readability formulas - even the one I recently perpetrated
myself [16] - do not predict readability. Those formulas which have been adequately
validated actually predict comprehensibility. Obviously, anyone who wants to go
on reading certain material must be able to understand it. On the other hand,
the fact that certain material is comprehensible to a certain person by no means
guarantees that he will find it readable.

I have therefore analyzed the reading habits of a stratified sample of
17,000 people and made statistical counts on a corpus of half a million words
culled from 47 newspapers and magazines in order to validate a sublimely simple
procedure that really does predict readability. The procedure is called
SMOG Counting. The name SMOG is, of course, a complimentary allusion to
Gunning's Fog Index, and to my birthplace, London, where smog originated, although
it has since been improved upon in several American cities. SMOG also happens
to be an acronym for Simple Measure Of Gobbledygook.

So that you will not be disappointed later on, let me disappoint you now.
Improved data validated for American readers should be published within a year,
but the SMOG Counts that I am presenting here have been standardised only for
British people, and they seriously underestimate the masochism of intelligent
readers.

However, let me, with characteristic British modesty and understatement,
demonstrate the virtues of SMOG Counting by contrasting it with the well-known [3]
Reading Ease formula devised by Rudolph Flesch.

Flesch's formula works like this. Take a sample of the prose you wish to assess. Determine the average number of syllables per hundred words, and the average number of words per sentence. Multiply the mean sentence length by 1.015 and the mean word length by 0.846. Add the two products together, and subtract the sum from 206.835. The result of all this numerology is a figure which, to the nearest whole number, is claimed to correspond to the school grade level which a reader must have attained if he is to understand the prose you have sampled.

It may well be objected that this formula ignores many factors which obviously go to determining the difficulty of a piece of prose. Quite apart from its legibility or audibility, the reader's or listener's comprehension will be affected by his interest in the topic and the amount of knowledge he can bring to bear upon it, by the logical coherence of the exposition, and by the ratio of its number of words to the number of ideas presented. The formula ignores all these factors, because they are far more difficult to measure than word and sentence length. This does not mean, however, that Flesch's formula is necessarily a poor predictor of comprehension.

It is not essential that the variables in a prediction formula should have direct causal connection with the quantity that is being predicted. All that matters is that these variables should correlate with the quantity to be predicted. Thus if we found that incompetent journalists were healthy, clean-living people, but that good journalists had ulcers, bad sight, smoked like chimneys and drank like fish, a formula based on measures of health and habits might predict a person's likelihood of succeeding in journalism far better than one based on measures with greater face value, such as verbal fluency and swift thinking.

The only criterion of a prediction formula is, then, that it should predict accurately. To find out whether this is the case for a particular formula

one must perform a validation study. If the criterion of readability is taken
to be mere comprehension, then validating a readability formula entails
measuring the word and sentence lengths of various prose samples, getting some
people educated to various grade levels to read these samples, determining
whether or not they understand the materials, and comparing their actual powers
of comprehension with those predicted. This sounds simple but in fact it verges
on the impossible. The main trouble is that there is no wholly satisfactory
way of measuring comprehension, as can be seen if we review the four main kinds
of comprehension test.

1. An _activity_ test of comprehension consists in asking a subject to
carry out some activity well within his grasp, according to certain instructions.
If he succeeds he has presumably understood the instructions. This kind of test
is obviously limited to instructional material.

2. A _replacement_ test requires subjects to replace words deleted from a
text: however this Cloze procedure tests a subject's understanding of the
mutilated text, which may well demand a kind of comprehension different from that
required to grasp normal, intact prose.

3. A _paraphrase_ test requires subjects to restate or summarise a text in
their own words. This has many shortcomings: a subject may well understand
what he reads but be unable to paraphrase it because his powers of expression
are too limited; success in the test merely shows that the subject can translate,
not that he has necessarily grasped the significance of the text; and, above all,
there is the difficulty of scoring the paraphrases: if the scoring is done
objectively, then this generally boils down to a test of ability to recall or
recognize certain crucial items: if one tries to get round this by asking the
subject to choose between a number of given summaries, then it becomes a test of
his ability to understand a summary, which is almost certain to be more difficult

than the original text; and if you try instead to get qualified judges to assess the subject's own summaries, you will find that their judgments are heavily affected by the mere length of the summary.

4. Lastly, one may try a question test, that is to say one asks questions about the text. This raises all the problems associated with a paraphrase test and a few more as well. To what extent are the answers implicit in the questions? Can a subject answer them anyhow, without having understood the text? Are the questions more difficult than the text? To what extent do the questions test reasoning and memory, rather than comprehension?

That last query pinpoints the difficulty of trying to predict comprehensibility: it shows that the concept of comprehension is not clearly definable. Nonetheless, a combination of the various alleged tests of comprehension could at least give us some idea of the validity of a comprehension prediction formula.

The real shock comes when one searches the literature for validation studies. Considering that Flesch's measure of Reading Ease and another similar formula, Gunning's Fog Index, are widely quoted in textbooks on journalism and technical writing as being adequate predictors of comprehensibility for adults, it is staggering to find that for the Flesch formula there are only half a dozen validation studies based on measures of adult comprehension, and none at all for Gunning's Fog Index. Even such studies as there are of the Flesch formula are not very convincing. One study does report a correlation of 0.87 between predicted reading ease and the percentage of parents able to answer comprehension questions; they were tested on 16 500-word samples drawn from magazine articles on parent health education.[10] The same study reports a 0.55 correlation between predicted reading ease and the comprehension of very poor adult readers; they were tested by being asked to select from five alternatives the best summary of

each of 40 100-word samples, drawn from reading material intended for poor adult readers, and also to select a detail not in the sample. The percentage of correct answers given by a cross-section of the British population to specific questions on 26 5-minute broadcast talks presented under good conditions did not correlate significantly with predicted reading ease. And the remaining three studies were on such a small scale that they could report only a positive relationship between predicted and observed comprehensibility among adults.[5,11,12]

This is not to say that the Flesch formula is useless. In fact it predicts the comprehensibility of material intended for children with a standard error of only 0.05 of a grade. Roughly speaking, this means to say that only 5 per cent of children who can correctly answer half the questions on a sample of prose will be more than two grades lower in school than the grade predicted for that sample.

Now I have let the cat out of the bag. What everybody who applies the Flesch formula to adult reading material chooses to overlook is the fact that it was calculated from children's comprehension scores, as given in "Standard test lessons in reading" by McCall and Crabbs.[14]

Flesch blandly assumed that if a child needs to have attained a certain reading grade to be able to understand a given passage of prose, then that is the school grade which an adult must have reached by the end of his education if he is to understand the same passage.

Quite apart from the obvious undesirability of calculating a formula to assess adult understanding on measures derived from material selected for children, there is a particular objection to using test lessons: doubtless the material was selected by McCall and Crabbs on account of the degree of difficulty they judged it to have, and these judgments of difficulty may well have been based to some extent upon their impressions of the word and sentence lengths of the materials.

Furthermore, there are deficiencies in the method of standardisation used by McCall and Crabbs.

Flesch also assumed, without producing any justificatory evidence, that semantic and syntactic difficulty are additively related to comprehension. For instance, according to the Reading Ease formula, if an author changes his style by halving the length of his sentences, his writings will increase in comprehensibility by the same amount whether he uses a simple vocabulary or one replete with sesquipedalian verbiage the meaning of which would remain esoteric even when set in the shortest of sentences.

By this time you perhaps have guessed that I am not particularly awe-struck by the Flesch yardstick of Reading Ease. But what about Gunning's Fog Index?[7] It has the merit of simplicity. Gunning proposed that the final reading grade an adult requires to understand a magazine article could be calculated by adding the average sentence length to the percentage of words of three or more syllables and multiplying this sum by two-fifths. And how did Gunning validate this inspired formula? By noting that when applied to some popular adult magazines it gave predictions which nicely accorded with his private judgments and - guess what - the grades required for comprehension as predicted by the Flesch formula.

Even if so-called readability formulas were properly validated on measures of comprehension they would still not predict readability, which, as I argued at the beginning of this paper, is a matter of willingness to read material rather than ability to comprehend it. In practice an editor is not generally worried about whether a potential reader can understand a story; what matters is whether the fellow will even try to read it beyond the first couple of paragraphs.

In order to find out what does determine readership I obtained data on the reading tastes of 17,600 informants representative of the entire British population aged 16 and above during the latter part of 1965 and early 1966.

This data was derived from a valid and reliable continuing survey of newspaper
and magazine readership carried out by the Institute of Practioners in
Advertising - IPA for short.[9]

The informants were classified on four variables: sex, age, class and
educational level. Class is closely correlated with educational level, but
earlier studies suggest that terminal educational age - the age at which a person
finishes his full-time education - is an important determinant of readership in
its own right.

To ensure that there was an adequate number of informants in each category
of my cross-classification, I divided them into only two age groups: those
aged 44 or less, and those aged 45 or more. I also divided them into only two
sexes - the IPA distinguishes three: men, women and housewives! The informants
were divided into four socio-economic classes: the upper and middle class,
designated AB; the C1 class, which includes supervisory and clerical workers; the
C2 class, which includes the skilled workers; and the DE class, which includes
more or less unskilled workers and those at the lowest levels of subsistence.
Three terminal educational age groups were distinguished: 15-, meaning those
who left school at the age of 15 or less; 19+, meaning those who continued in
full-time education until the age of 19 or more; and the intermediate 16 through
18 age group. Only two adjustments to the data appeared necessary: informants
still in full-time education at the age of 18 were included in the 19+ group:
and the 200 people whose terminal educational age was not stated were excluded
from the study, but as most of them probably belonged to the 15- group, like the
vast majority of other informants, this exclusion is unlikely to have biased the
results at all seriously,

For 47 varied newspapers and magazines I thus obtained readership figures
in 48 categories, each classified by sex, age, class and educational level,

Correlations between readers of different ages or different sex within a given

class and educational group were all of the order of 0.9. It was therefore

justifiable to make the readership data more amenable to factor analysis by

conflating the age and sex variables. Because there were very few people in

in the C1, C2, and DE classes with an educational level of 19+, I also collapsed

these three groups into one. This reduced the original 48 categories to 10.

A principle components analysis of these readership figures showed that

they can be explained by a delightfully simple model. No less then 97 per cent

of the total variance in the readership of the 47 periodicals is accounted for

by two components. However, 77 per cent of the variance is due to a factor

clearly identifiable as 'general interest: periodicals scoring high on this

component appeal to all classes and educational levels of men and women,

periodicals with low scores on this component have a much more limited appeal.

The other factor is clearly identifiable with linguistic difficulty. When

periodicals are ranked by scores on this factor, the rankings correlate 0.57 with

rankings by sentence length, and 0.56 with rankings by word length. Those

correlations are significant at the 1 per cent level and are high enough to

justify identifying the factor with linguistic difficulty.

Independent studies by Coleman[2] and myself[15] show that the lengths of

sentences have much less effect upon comprehension than the lengths of their

constituent clauses. I have demonstrated that when other words - such as the

ones which I am inserting at this very moment into an otherwise fairly

straightforward sentence - intervene between grammatically related words, this

constitutes a prime source of difficulty for a reader or listener.

My theory to account for the psychological difficulty induced by even

small amounts of separation of grammatically related words is this: assume that

the process of perceiving a segment of a sentence sets upon some kind of pattern

of activity in the brain; assume also that this pattern of activity decays rather rapidly, but that comprehension depends upon one combining patterns evoked by grammatically related elements, so that they have to be retained simultaneously. It follows that the greater the separation between two related segments, the greater is the probability that the pattern evoked by the first segment will have decayed beyond recall before the second is perceived, so that the probability of complete comprehension of the sentence in which the segments occur is reduced. This hypothesis is supported by a recent finding that increasing separation merely by replacing some words by familiar but longer equivalents will increase the difficulty of a sentence.

Likewise, word length itself does not make for difficulty in comprehension: but, by an historical accident, the everyday words in our language have Anglo-Saxon roots, which tend to be short, whereas learned words come from classical languages which rejoiced in polysyllabic vocables. These longer words are used only when greater precision is required. I hypothesize that identifying the meaning of a word involves categorising it; the more precise the word the more categorisations are required before it is identified. But the longer it takes to locate a word's meaning, the more likely it is that the preceding context will be lost beyond recall. Thus word length, like sentence length, is an index of difficulty due to limitations of immediate memory.

To check that difficulty is indeed related to vocabulary precision I counted the length in letters of 10,000 word tokens in matched samples from three British newspapers, The Times, the very popular Daily Mirror, and the Daily Mail, a mass-circulation paper of sligthly greater difficulty than the Mirror.

A chi-squared test showed that there was no significant difference in the frequency distribution of words in the two papers. Furthermore, use of a formula

derived by Guiraud[6] suggested that the size of the vocabulary drawn upon by Mirror writers was about 20,000 words - and that the Times writers were using a vocabulary only very little larger. Of course, the vocabulary of Times writers is not the same as that of their colleagues on the Mirror: The Times contains many more long, unfamiliar but more precise words. This difference was clearly reflected in differences in the proportions of words of seven or more letters long: there were 249 of these per thousand in the Times, compared with 227 in the Daily Mail, and 219 in the Mirror.

Thus there is both empirical and theoretical justification for using measures of word and sentence length as predictors of readability. But there are more ways than one of measuring these lengths.

The frequency distribution of both word lengths and sentence lengths are severly skewed. Believing that the occasional extremely long words and sentences are responsible for most of the difficulties experienced by readers, I decided to examine the extremes of the distributions rather than their means. I therefore measured sentence length by the percentage of sentences in a sample which were more than 20 words long; and word length by the percentage of polysyllables, that is the percentage of words in a sample which were three or more syllables long.

Interestingly enough these measures provide just about as much information as the much more tedious calculations of mean word length and mean sentence length. From data yielded by the Brown University count of a million words I[13] have derived a regression formula which relates mean sentence length to the percentage of long sentences, that is, of more than 20 words. This formula has a multiple correlation coefficient of .995, but a reasonable approximation to mean sentence length is given simply by the percentage of long sentences multiplied by 3.

It is my belief that word lengths have a negative binomial distribution.
If this speculation is correct then there must be a very close relationship
between the mean word length of a sample and the percentage of polysyllables it
contains. However, I find that a good approximation to the number of syllables
in a hundred words is given by adding 113 to the percentage of polysyllables
multiplied by 3.

Admittedly, syllable measures of word length are not entirely reliable
because not only do our ideolects differ but so do our intuitions about syllabic
structure - take that word "our", does it contain one, two or three syllables?
The trouble is that a syllable is a phonetic concept, not an orthographic one:
syllables can be defined rigorously only in terms of chest pulses which cannot
be counted without special apparatus, and it must be remembered that syllables
are modified by rate of speech. Greater reliability can be obtained by counting
individual speech sounds (phonemes) or, better still, individual letters: this
latter task can even be performed by computer.

Yet it seems that whether you count syllables, phonemes or letters, you
get just about the same amount of information. From analysing data on telephone
conversations which included more than 76,000 word tokens,[4] I find that the
average number of letters in a word, whether long or short, consistently averages
1.23 times the number of its phonemes. Furthermore, the number of syllables is
approximately one-third of the number of letters in a sample.

So it seems that a count of polysyllables is at least as valid as any
other index of word difficulty, and it is certainly the easiest count to make.
Of course if the number of polysyllables is to be expressed as a percentage of
the words in a sample, the chore of counting off at least one hundred words
cannot be avoided. There is, however, a still easier alternative. Simply count
off ten sentences, marked by periods, question marks or exclamation marks:

then count the number of polysyllables in those ten sentences. This measure, which is called a SMOG Count, is equivalent to multiplying the percentage of polysyllables by one-tenth of the mean sentence length. Thus with the minimum of effort you can obtain a measure of both semantic and syntactic difficulty combined by making a SMOG Count.

Because word and sentence lengths are locally variable it is wise to make three counts and average them. Haskins reports that the linguistic characteristics of samples taken from the start, middle and end of magazine items when combined correlate .95 with characteristics of the entire item.[8]

A complication now arises because there is a tendency for samples taken near the beginning of any story to have greater linguistic difficulty than the remainder of the story. The average percentage of polysyllables near the start of articles is 10.73 compared with 10.17 towards the end; the average percentage of sentences more than 20 words long is 35.24 near the beginning compared with 33.14 later. This means that, so far as linguistic difficulty is concerned, the crucial part of an article is generally near the beginning, and that is where SMOG Counts should be made.

To discover the average level of word and sentence difficulty of the reading material preferred by a typical member of each of the original 46 categories of reader, the following procedure was adopted.

First the periodicals read by the 17,400 informants were sampled, using a method based on the notion that a reader does not read every word in a periodical, but rather that he selects those articles which promise to interest him most. Three large panels of readers were therefore set up for each periodical. The panels consisted of university students aged between 17 and 27, each of whom had stated a preference for reading one of the 47 periodicals included in the study. Each panellist was given 15 sets of sheets taken randomly from issues

of his or her periodical published at the end of the period covered by the IPA survey. Each set of sheets was extensive enough for the panellist to find an article not less than 300 words long which interested him. In each such article the panellist then marked the starts of two passages for analysis: one mark was to be made one-fifth of the way through the article, the other three-fifths of the way through as judged by eye.

The marked sheets were then collected and redistributed at random to a large group of volunteer analysts. Each analyst was asked to count 100 words starting at each mark, noting the number of polysyllables within those 100 words. He was also asked to count in each passage five sentences, noting the number of words in each sentence.

Thus 45 pairs of passages of interest to a more or less typical reader were analysed for each periodical. A random subsample of passages was reanalysed by a second set of analysts, and where their results indicated that one of the original analysts had been grossly careless, that person's analyses were discarded. This was not the best procedure, but it was the only practicable one considering that nearly half a villion words were being counted. As many as three pairs of passages were ignored in some periodicals, but for the majority only one pair had to be excluded.

Weighted means for the linguistic difficulty of typical reading matter chosen by each of the 48 categories of reader were then calculated in this way; the average percentage of polysyllables in samples drawn from the first halves of articles in each periodical was multiplied by the number of respondents who read that periodical; the products for all 47 periodicals were then added together: finally, this sum was divided by the total number of responses. The same process was used to obtain weighted means for sentence length. The resulting figures, I repeat, gave two average measures of linguistic difficulty for the

periodical reading material chosen by typical people in each of 48 categories.

One interesting finding that then emerged was that the measures of word and sentence length have a correlation of .91. This is presumptive evidence for my contention that semantic and syntactic difficulty are both related to limitations of short-term memory storage. Although his manuscript helps to extend his memory, a writer still has to remember what he is trying to put into a sentence while he is composing it.

The calculations of weighted means for word and sentence length were replicated with samples drawn from the latter halves of articles. This made it possible to carry out analyses of variance, treating the replication factor and its interactions as the residual error term. The analyses indicate that a person's educational level, socio-economic class, sex and age are highly significant in determining the degree of linguistic difficulty with which he is willing to cope in this reading matter. What is really surprising is that for all practical purposes these variables do not interact. That is to say, high educational level will incline a person to read material of x more units of difficulty than material which a person of low educational level would choose; high class will incline him to read material of y more units of difficulty than someone of low class would choose; but if he is both of high educational level and high class he will tend to choose material x plus y units more difficult than an uneducated low-class person would select. The variables do not interact; they are simply additive.

The implications of this finding became clear as soon as preferred SMOG Counts for each of the 48 categories of reader were computed. This was done by using the regression formula previously mentioned to transform the sentence measures to mean sentence lengths, multiplying these by the percentage of polysyllables divided by ten, and rounding off the results to the nearest

whole number. By comparing an actual SMOG Count with a table of preferred Counts it is easy to see which of the 48 groups of readers would like the material sampled, which groups would find it too hard, and which too simple. Tables of preferred average word and sentence lengths can be used similarly, but, of course, SMOG Counts take less trouble.

Because the factors determining readers' linguistic preferences are additive, one need not even trouble to look up the table of preferred SMOG Counts. The preferred Count for any given group of readers can be estimated simply by adding weights to the preferred Count for reading matter chosen by typical uneducated girls of the lowest social class. This base is 15. Make the indicated addition for each of the following conditions which apply to the given group.

    Add 1 for: age above 44
                 belonging to the C1 social class
                 full-time education finished at 16, 17 or 18

    Add 2 for: men
                 belonging to the AB social class
                 full-time education continued beyond 18

This system of weights gives the correct preferred average SMOG Count for 30 of the 48 categories; it is only one unit out for 14 further categories.

Remember that the preferred SMOG Counts are standardised only for British readers. Differences in educational and class structure probably make these standards inapplicable elsewhere. Furthermore, the preferred Counts are based only on data for large-circulation newspapers and magazines. Other, more specialised reading matter, such as professional journals, have literally been left out of the count. But perhaps it was worth while doing this preliminary study without reference to specialised journals, if only to make this final observation. Preferred SMOG Counts range from 15 up to only 25 for the most serious, best educated readers of the highest social class. Yet anyone who cares to apply a SMOG Count to a professional, technical or learned publication will

find that Counts of 100 or more are quite common. Can we really say that we are witnessing an information explosion if all that is happening is that the sum of human gobbledygook doubles every decade?

Analysis of variance of percentages of polysyllables
in reading material chosen by
48 categories of reader

Source of variation	Degrees of freedom	Mean squares	F
Education (E)	2	5.52733	45.421***
Class (C)	3	3.12396	25.671***
Sex (S)	1	8.80275	66.421***
Age (A)	1	2.50194	20.560***
E x C	6	0.32689	2.686**
E x S	2	0.60365	4.961**
E x A	2	0.04579	0.376
C x S	3	0.39724	3.264**
C x A	3	0.02335	0.192
S x A	1	0.01042	0.086
E x C x S	6	0.33170	2.726**
E x C x A	6	0.06481	0.533
E x S x A	2	0.14619	1.201
C x S x A	3	0.10018	0.823
E x C x S x A	6	0.17647	1.450
Error	48	0.12169	

*** Significant at 0.1% level

** Significant at 5% level

Analysis of variance of percentages of sentences more than 20 words long in reading material chosen by 48 categories of reader

Source of variation	Degrees of freedom	Mean squares	F
Education (E)	2	69.29651	13.6580***
Class (C)	3	45.07303	8.8837***
Sex (S)	1	113.05737	22.2831***
Age (A)	1	15.39205	3.0337*
E x C	6	2.76608	.5451
E x S	2	3.14195	.6192
E x A	2	0.35429	.0698
C x S	3	1.26811	.2499
C x A	3	0.44521	.0877
S x A	1	0.36506	.0719
E x C x S	6	1.20122	.2467
E x C x A	6	1.20492	.2374
E x S x A	2	1.85317	.3652
C x S x A	3	0.40641	.0801
E x C x S x A	6	1.57620	.3106
Error	48	5.07369	

*** Significant at 0.1% level

* Significant at 10% level

Average percentages of polysyllables
in reading material chosen by
48 categories of reader

Socio-economic class	Terminal Educational Age		
	19+	16 - 18	15-
AB	13.23 11.89 11.70 11.67	11.43 11.34 11.24 11.01	11.12 10.72 10.71 10.38
C1	12.80 11.25 11.34 11.00	11.35 11.03 11.17 10.56	10.90 10.75 10.53 10.19
C2	12.65 11.70 10.70 10.38	10.91 10.78 10.48 10.09	10.54 10.39 10.17 9.91
DE	12.00 11.26 10.56 10.09	10.80 10.73 10.40 10.05	10.40 10.34 10.04 9.69

Data for four groups of people are given in each main cell according to the following scheme:

Men aged 45+

Men aged 44-

Women aged 45+

Women aged 44-

Based on about 44 100-word samples (taken near the start of articles) from each of 47 British newspapers and magazines.

Average SMOG Counts of reading
material chosen by
48 categories of reader

Socio-economic class	Terminal Educational Age		
	19+	16 - 18	15-
AB	25 21 20 20	20 20 19 19	19 18 18 17
C1	24 20 19 19	20 19 19 17	18 18 18 16
C2	23 21 18 17	19 18 17 16	17 17 16 16
DE	20 19 17 16	19 18 17 16	17 17 16 15

Calculated from: $1/10 \times P \times S$

where   P = mean percentage of polysyllables

S = mean sentence length

= $12.87848 + 0.46312\ SS - 2.02755 \sqrt{SS}$

and   SS = mean percentage of sentences more than
20 words long

Data for four groups of people are given in each main cell according to the following scheme:

Men aged 45+
Men aged 44-
Women aged 45+
Women aged 44-

Average percentages of sentences
more than 20 words long in
reading material chosen by
48 categories of reader

Socio-economic class	Terminal Educational Age		
	19+	16 - 18	15-
AB	41.24 37.86 35.92 35.00	35.20 36.06 35.36 34.28	34.98 34.06 32.70 31.54
C1	40.00 35.84 34.46 34.72	35.68 35.38 33.86 32.62	33.98 34.10 35.16 31.40
C2	37.50 36.60 33.20 32.02	34.72 34.22 32.32 30.78	32.94 32.74 31.18 30.32
DE	33.22 34.94 32.98 32.16	35.40 33.52 33.12 30.84	32.50 32.24 30.62 29.76

Data for four groups of people are given in each main cell according to the following scheme:

Men aged 45+
Men aged 44-
Women aged 45+
Women aged 44-

Based on about 44 5-sentence samples (taken near the start of articles) from each of 47 British newspapers and magazines.

Linguistic characteristics of British newspapers and magazines

Periodical	% Sentences over 20 words long Mean	S.E.	% Polysyllables Mean	S.E.
Financial Times (D)	58.87	8.48	16.59	0.58
The Times (D)	54.44	3.16	15.28	.60
Do It Yourself (MM)	51.17	2.61	8.54	.50
Car Mechanics (MM)	49.31	2.72	8.49	.40
Guardian (D)	47.20	3.69	13.84	.56
Practical Householder (MM)	46.67	2.72	7.89	.45
Radio Times (W)	44.89	3.72	13.50	.66
Sunday Telegraph (S)	44.78	2.92	13.15	.57
Practical Motorist (MM)	44.66	2.64	10.57	.57
Daily Telegraph (D)	44.10	2.58	15.38	.59
Sunday Times (S)	42.73	3.00	13.66	.54
Punch (W)	42.40	3.12	11.75	.52
Observer (S)	40.47	3.02	12.68	.52
Evening Standard (D)	38.84	2.66	12.42	.55
The Universe (W)	38.61	2.98	13.43	.68
Ideal Home (MW)	37.99	2.46	12.37	.49
Evening News (D)	36.88	2.92	12.38	.59
Daily Express (D)	36.43	2.96	11.12	.54
The Sun (D)	36.37	1.96	11.89	.48
Christian Herald (W)	36.37	2.80	8.42	.55
Sunday Mail (S)	34.00	2.92	9.22	.50
Good Housekeeping (MW)	33.18	3.56	10.69	.49
News of the World (S)	32.80	3.32	9.62	.64
Vogue (MW)	31.56	2.68	12.34	.61
Sunday Express (S)	31.40	2.58	10.11	.56
Daily Sketch (D)	31.40	2.78	11.65	.58
Daily Mail (D)	31.37	2.36	11.84	.47
Sunday Citizen (S)	31.14	2.92	13.01	.62
Woman's Mirror (WW)	28.86	2.38	9.55	.54
Honey (MM)	28.45	2.68	9.18	.53
Sunday Mirror (S)	28.22	2.44	11.14	.59
Woman's Own (WW)	28.00	2.24	7.87	.34
She (MM)	28.00	2.88	10.08	.54
Daily Mirror (D)	27.05	2.50	10.85	.64
T.V. Times (W)	26.19	2.26	8.79	.55
Woman's Realm (WW)	25.91	2.40	8.32	.39
Reader's Digest (M)	25.69	2.28	10.77	.60
The People (S)	25.33	2.44	9.22	.45
Sunday Post (S)	25.23	2.84	9.59	.53
Reveille (W)	25.00	2.32	8.49	.58
Parade (W)	24.77	2.64	8.76	.54
Week-End (W)	24.55	2.46	8.59	.57
Woman's Weekly (WW)	24.54	2.54	7.07	.38
Woman (WW)	22.22	2.34	7.73	.40
Tit-Bits (W)	21.55	2.38	8.59	.44
Valentine (WT)	14.22	2.14	10.08	.54
True Stories (WT)	10.64	2.08	4.48	.28

D Daily newspaper
S Sunday newspaper
W Weekly magazine   WW for women   WT for teenage girls
M Monthly magazine   MW for women   MM for men

Based on about 88 100-word samples for each periodical.

REFERENCES

1. BELSON, W.A., 1952, _An enquiry into the comprehensibility of 'Topic for tonight'_, Listener research report LR/52/1080, Audience Research Dept., BBC, London.

2. COLEMAN, E.B., 1962, Improving comprehensibility by shortening sentences, _Journal of Applied Psychology_, 46, 131-134.

3. FLESCH, R.F., 1948, A new readability yardstick, _Journal of Applied Psychology_ 32, 221-233.

4. FRENCH, N.R., CARTER, C.W., and KOENIG, W., 1930, The words and sounds of telephone conversations, _Bell System Tech. J._, 9, 290-324.

5. GRIFFIN, P.F., 1949, Reader comprehension of news stories: a preliminary study, _Journalism Quarterly_, 26, 389-396.

6. GUIRAUD, P., 1954, _Les caracteres statistiques du vocabulaire_, Presses Universitaires de France, Paris.

7. GUNNING, R., 1952, _The technique of clear writing_, McGraw-Hill, N.Y.

8. HASKINS, J.B., 1960, Validation of the Abstraction Index as a tool for content-effects analysis and content analysis. _Journal of Applied Psychology_, 44, 102-106.

9. INSTITUTE OF PRACTITIONERS IN ADVERTISING. Raw data, available only in punched card form, collated for this study by Research Services, London.

10. KLARE, G.R., 1952, Measures of the readability of written communication: an evaluation, _Journal of Educational Psychology_, 42, 385-399.

11. KLARE, G.R., MABRY, J.E. and GUSTAFSON, L.M., 1955, The relationship of style difficulty to immediate retention and to acceptability of technical material, _Journal of Educational Psychology_, 46, 287-295.

12. KLARE, G.R., SHUFORD, E.H., and NICHOLS, W.H., 1957, The relationship of style difficulty, practice, and ability to efficiency of reading and retention, _Journal of Applied Psychology_, 41, 222-225.

13. KUCERA, H., and FRANCIS, W.N., 1967, Computational analysis of present-day American English. Providence, R.I.: Brown University Press. Tables D2 and D12-26.

14. McCALL, W.A., and CRABBS, S.M., 1925, _Standard test lessons in reading_, Teachers College, Columbia University, New York.

15. McLAUGHLIN, G.H., 1966, What makes prose understandable: an experimental investigation into comprehension. Ph.D. Thesis, University College London.

16. McLAUGHLIN, G.H., 1969, SMOG Grading - a new readability formula, _Journal of Reading_, 639-646.